PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : G06F 15/16

A1

(11) International Publication Number: WO 00/42519

(43) International Publication Date: 20 July 2000 (20.07.00)

(21) International Application Number: PCT/US00/00587

(22) International Filing Date: 11 January 2000 (11.01.00)

(30) Priority Data:

60/115,456	11 January 1999 (11.01.99)	US
60/169,971	8 December 1999 (08.12.99)	US
Not furnished	10 January 2000 (10.01.00)	US

(71) Applicant: EDGIX CORPORATION [US/US]; 1322 East 550 North, Orem, UT 84097 (US).

(72) Inventors: POWELL, Kyle; 1322 East 550 North, Orem, UT 84097 (US). HURST, Mark; 6222 West 10480 North, Highland, UT 84003 (US). MERKLEY, Karl; 1104 East 230 North, Lindon, UT 84042 (US).

(74) Agent: WITT, Evan; Madson & Metcalf, 15 West South Temple, Suite 900, Salt Lake City, Ut 84101 (US).

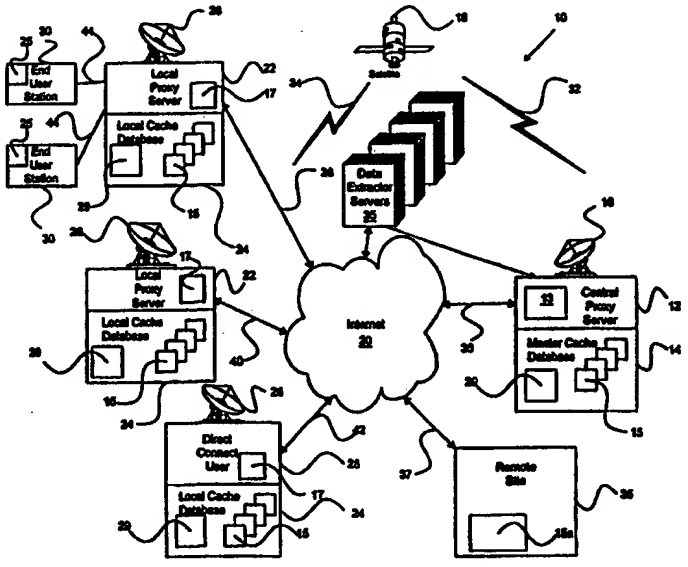
(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: INTERNET CONTENT DELIVERY ACCELERATION SYSTEM



(57) Abstract

A system and method accelerates distribution of content of a global communications network such as the Internet (20). A central proxy server (12) transmits extracted data files over a communications medium to provide content filling of local proxy servers (17). The communications medium may be satellite transmission and may be conducted using IP multicast. The local proxy servers concurrently receive the data files from the communication medium at a high rate of speed and store the data files in attendant local cache databases (29). The local proxy servers utilize a localized heuristics scheme to determine whether to keep or discard the data. When a user requests Internet data, the user's request is first received by the local proxy server. If the requested data is present among the extracted data files, it is rapidly transmitted directly to the user from the local proxy server, if not present, a request is transmitted to the central proxy server.

INTERNET CONTENT DELIVERY ACCELERATION SYSTEM

BACKGROUND OF THE INVENTION

1. The Field of the Invention

The present invention relates to Internet acceleration systems. More specifically, the present invention relates to Internet Acceleration systems involving local and remote caching to speed up distribution of content on the Internet.

2. The Relevant Art

Gridlock has hit the Internet--too much Internet traffic and too little bandwidth are resulting in slow response times. One solution being developed in response to this problem is caching. It is estimated that 60 to 80 percent of Internet traffic is cachable. Taking the web pages hit most often, and locating them near the edge of the network and close to users at the business or ISP site dramatically reduces Internet traffic. With caching solutions, storage is used to improve Internet performance instead of adding communication bandwidth. This has important implications because the cost of storage continues to fall while the cost of bandwidth remains relatively expensive.

While caching solutions can significantly boost Internet performance, two major problems remain. First, while backbone operators that sell directly to large customers have sufficient "critical mass" to benefit significantly from caching, businesses and smaller ISPs do not. The success of web page caches is a function of "hit rate" (percentage of requests where data is already in cache). But there is an enormous amount of web pages available and in a small population of varied end users there is a much smaller probability that the request of a particular end user will already be cached than with large populations of end users. The second problem is that in order to fill the cache, web pages still transit the Internet backbone, just not as often. With current caching systems, only subsequent references to web pages avoid the Internet connection and have faster access times.

Solutions for improving caching systems are needed. Among those solutions, improved systems for increasing hit rates within caches would be a great improvement in the art. Additionally, improved manners of extracting web data and distributing that data to local caches would be very helpful. Particularly helpful would be manners of communicating with local caches to determine which web pages are most popular and transmitting that popularity information to a central location where it can be compiled and used to select web data for

module in integral communication with the cache database management module. Preferably, the integral communication comprises direct communication between programs on a common server or other computer, and may be facilitated by an applications program interface (API).

Because of the tight integration between the cache database management module and the local caching module, sophisticated heuristics determinations may be employed. For instance, rather than simply registering and reporting to the central proxy server when a data file is requested and is not present in a local cache, customized hit information for all transmitted data files and even data files that have not been transmitted may also received from the caching databases and transmitted to the central proxy server for analysis and use in determining which data files to obtain and forward to the local proxy servers. Additional determinations may be made regarding characteristics such as the timing of demand for data files, together with the nature of the data file for customized determinations of demand within the individual local proxy servers.

When a user requests Internet data, the request is first sent to the local cache database management module to determine whether the requested data is present. If the data is present, it is rapidly transmitted to the user from the local cache database, preferably avoiding any transmission over the Internet. If the data is not present, a request is transmitted to the central proxy server which requests a copy from the hosting site.

The system may also transmit selected data to subscribing local intranet sites to rapidly and systematically publish private data to local proxy servers connected to the local intranets.

BRIEF DESCRIPTION OF THE DRAWINGS

In order that the manner in which the advantages and objects of the invention are obtained will be readily understood, a more particular description of the invention briefly described above will be rendered by reference to specific embodiments thereof which are illustrated in the appended drawings. Understanding that these drawings depict only typical embodiments of the invention and are not therefore to be considered to be limiting of its scope, the invention will be described and explained with additional specificity and detail through the use of the accompanying drawings in which:

Figure 1 is a schematic block diagram illustrating one embodiment of an Internet content delivery acceleration system of the present invention, including a central proxy server and a plurality of remote proxy servers.

servers 22, each having a local caching module 17 disposed therein. Each local proxy server 22 also comprises or is otherwise in communication with a local cache database 24. The local proxy servers 22 in one embodiment are high-performance, high-capacity PC-based servers. In one embodiment, the proxy servers operate under an IPX protocol, such as Novell's Netware™. Facilities in which the local proxy servers 22 may be installed include Internet service providers (ISPs), large and medium businesses, and eventually, as economies of scale make the service highly affordable, small businesses and homes.

The local proxy servers 22 are each provided with a local cache database 24. Preferably, the local cache databases 24 are provided with high capacity memory. In one embodiment, the cache databases 24 are provided with hard drive memory in excess of 40 Gigabytes. A caching database management module 29 is provided to interface with the local cache database 24. In one embodiment, the caching database management module comprises an Internet caching system (ICS) program 29, such as Novell Corporation's Border Manager™ and/or ICS™ software. Preferably, the local proxy servers are interfaced in integral communication with the caching database management modules, as will be described in greater detail below.

The local proxy servers 22 preferably provide users at end stations 30 with access to a global communications network, such as the Worldwide Web existing on the Internet 20. This makes them ideally suited for installation at locations such as ISPs, telephone system switching backbones, or points of presence (POPs), and within networks of large companies, such as fortune 500 companies. Thus, an end user at a station 30 may be connected to a local proxy server 22 remotely over a modem, or locally over a network 44. The local proxy servers 22 can be considered to be at the "edge" of the Internet 20, a position allowing the present system 10 to reduce traffic over the conventional transmission routes of the Internet 20.

To do so, the local proxy server servers 22 receive and store high-demand data files 15 in the attendant local cache databases 24. In an embodiment to be described herein, the network data comprises web pages 15a emanating from a remote Internet site 35. The web pages 15a are initially gathered by the central proxy server 12 from Internet sites 35 over a communication channel 36 and are subsequently transmitted from the central proxy server 12 to the local proxy servers 22. The transmission is conducted over a suitable communication medium.

If the requested web page 15a is not present in the local cache database 24, the http caching protocol causes the request to be passed on through the Internet 20 to the Internet site 35 at which the web page 15a is hosted. The Internet site 35 then transmits the web page 15a over the Internet 20 back to the requesting user at a station 30. This Internet transmission occurs over regular Internet communication channels 38, 40, 42.

Concurrently, the request is also passed to the central proxy server 12. Once the central proxy server 12 receives its copy of the request, it also requests a copy of the requested web page 15a from the hosting Internet site 35. Accordingly, the web page 15a is also transmitted over channels 36, 37 to the central proxy server 12. Once the web page 15a is received, the central proxy server 12 caches the web page 15a within the attendant master cache database 14. It may also transmit the web page 15a to the communicating local proxy servers through the communication medium if the web page 15a is found to have a sufficiently high priority. This transmission in one embodiment is satellite transmission. The web page 15a is transmitted 32 through a satellite uplink transmitter 16 to the satellite 18 at a speed of, for example, 25 Mega bytes per second (Mbps). The satellite 18 then transmits 34 the web page 15a to each of the subscribing local proxy servers 22, at a rate preferably of about 25 Mpbs.

The central proxy server 12 may filter the web page 15a before caching it. It may also keep track of the number of requests and transmit web pages 15a (or other data 15) only after a certain minimum number of requests have been logged for that web page 15a. Accordingly, the central proxy server 12 is in a position to be a "Nelson Ratings System" for the Internet 20, having centralized access to web site demand information.

Additionally, the central proxy server 12 preferably receives global popularity information about data files 15 from subscribing local proxy servers 22. This popularity information in a basic embodiment comprises miss data. Additionally, due to the unique configuration of the present invention, more sophisticated information may also be transmitted. This may include, for instance, hit information and timing information, as will be discussed in greater detail below.

The central proxy server may also utilize data 15 popularity information, data 15 characteristics, data associations, and other data useful to local proxy servers 22 in deciding which extracted data 15 to continue to store. Such information may be collected from the

individualized local policy and a heuristics program to determine whether to keep the transmitted web page 15a or discard the web page 15a.

Reference is now made to Figure 2 for a discussion of a further aspect of the present invention. In addition to multicasting generally to a series of local proxy servers 22, the central proxy server 12 may also service one or more virtual private intranets 50 through selective pointcast satellite transmission 74. Under this concept, and in an embodiment to be specifically described, the system 10 preferably operates in substantially the same manner as described above. Accordingly, the system 10 is provided with the central proxy server 12 and a series of local proxy servers 22 configured substantially in the manner described. In addition, the system 10 may also be provided with the private intranet 50. The private intranet 50 is shown communicating over the Internet 20 through an intranet proxy server 60. The intranet proxy server 60 may also function as a local proxy server 22, as discussed above.

As depicted, the private intranet 50 connects to end users at remote stations 52 through a communication channel 54. The private intranet 50 is connected with the intranet proxy server 60 through a communication channel 68. The Intranet proxy server is in communication with a satellite down-link receiver 62 through a communication channel 66 and with the Internet 20 with a communications channel 64. One or more of the aforementioned channels may maintain security with a firewall 78.

The system 10 may also be provided with a private intranet publisher 70 for generating or otherwise providing private data 55 to be communicated to users at stations 52 within the private intranet 50. In the depicted embodiment, the private intranet publisher 70 is a dedicated server computer communicating over the Internet 20 through a secure channel 72. The secure channel 72 is shown provided with a firewall 78.

The private intranet publisher 50 generates or collects the private data 55, which may be business, financial, or any other type of data, and transmits the private data 55 in web page form. The private data 55 is passed in a secure manner through the Internet 20 to the central proxy server 12. The central proxy server 12, through a satellite uplink transmitter 16, then uplinks 32 the private data 55 to the satellite 18. The satellite 18 then transmits the private data 55 to intranet proxy servers 60 of the subscribing virtual private intranet using a focused multicast 74. The same satellite 18 may also transmit 34 other data such as web pages15a to the regular local proxy servers 22.

procedures, routines, or other executable software code. Nevertheless, the modules could also be implemented with other types of programmable logic such as programmable logic arrays (PLAs), ASICs, or even logic circuits or other types of hardware.

In order to initialize the system 10 (of Figures 1 and 2) and enable communication links, the depicted components of a system initialization block 80 is employed. The modules of the system enablement block 80 may be contained as software code within the central proxy server 12, or may be distributed amongst the different hardware components of the system 10. Initially, the central proxy server 12 is initialized and brought on-line with a central server enablement module 82. This preferably includes enabling Internet communication over communication channel 36 and enabling satellite communication with the satellite uplink transmitter 16.

Subsequently, the local proxy servers 22 are enabled and brought on-line with a local proxy server enablement module 84. Typically, Internet communications are enabled through communication channels 38, 40, and 42. It is of note that the various local proxy servers 22 will generally not all be brought on-line at one time. Typically, the system 10 of the present invention will be provided commercially as a service to which customers subscribe. As new customers subscribe, they are provided with the local proxy server 22 and satellite receiver hardware 26 for an initial fee. A monthly subscription fee may be charged thereafter.

Satellite communications are enabled with a satellite initialization module 86. This may include establishing the communications link between the satellite 18 and the satellite uplink transmitter 16 and between the satellite 18 and the satellite receivers 26, 62. It may also include establishing a proper protocol for satellite communications. For the embodiment of Figure 2, satellite communications for the pointcast transmission must be established with one or more Internet proxy servers 60.

The intent of the system 10 is to fill the local cache databases 24 of each subscribing local proxy server 22 with the most highly requested data such that the majority of requests for web pages 15a may be serviced directly by the local proxy servers without having to go to the Internet 20. Accordingly, each new central proxy server that is brought on-line must be filled with the most highly requested web pages 15a. This is conducted at with an initial filling module 88 and may be accomplished in a variety of manners. In one contemplated embodiment, the central proxy server 12 transmits "old" data that has been previously transmitted in between transmitting "fresh" data that has only recently been requested. This

Additionally, if the uplinked data 15 contains private data 55, a pointcast data module 116 formats and pointcasts the private data 55.

An uplink transmitter block 120 of Figure 3 illustrates general functional modules operating within the satellite uplink transmitter 16 of Figures 1 and 2. The data 15 to be uplinked is formatted into a proper form and protocol for satellite transmission with a data formatting module 124. Header information is added to the packets to be uplinked with a header addition module 124. One simplified example of the structure of a packet 190 and header 192 is shown in Figure 4. The uplink transmission is conducted between the satellite uplink transmitter and the satellite 18 with the use of an uplink transmission module 126.

An Internet block 130 illustrates one example of the functional modules operating within the Internet 20 of Figures 1 and 2. Requests for data 15 are transmitted over the Internet 20 with a data request module 132. The request is typically the request generated by the Internet request module 104. Once the data 15 is requested, the Internet site 35 at which the data 15 is hosted is contacted with a site contacting module 134. A site map of the site, including the location and configuration of the web page 15a is transmitted from the web site 35 with a site map transmission module 136. Thereafter, the web page 15a is transmitted from the Internet site 35 to the central proxy server 12 with a data packet transmission module 138. The web page 15a and attendant site map may also be transmitted directly to the requesting local proxy server 22.

A local server block 150 illustrates one embodiment of the functional components operating within the local proxy server 22 of Figures 1 and 2. Preferably, these functional components are contained within the local caching module 17 of Figure 1. Within the local server block 150, a data request block 155 includes a data reception module 152. The data reception module 152 receives requests for Internet data 15 from the end users at the stations 30. Generally, this request is generated by the data request module 92. A search module 154 searches the local cache database 24 for the requested data 15. In one embodiment, the search module 154 comprises the proxy cache protocol employed within the http language.

If the requested data 15 is not present within the local cache database 24, a data request module 156 transmits a request for the data 15 to the central proxy server 12. This request is typically routed over one of the communication channels 38, 40, or 42, through the Internet 20, and through communication channel 36 to the central proxy server 11. The request may comprise the original request received from the user at a station 30 modified with the Internet

storage of the data 15 onto disk at the high rate of speeds specified, and/or Novell's Border Manager™ software which acts as a proxy cache manager and provides proxy content filling and firewall functions.

A heuristic determination block 170 illustrates the functional components of one embodiment of a heuristic formula of the present invention. A local statistics collection module 172 resident within each local proxy server 22 monitors the frequency of requests for each file of data 15 located within the local cache database 24. A global statistics reception module 174 receives data demand statistics from a central location. In one embodiment, the central location is the central proxy server 12, which collects demand statistics during the process of servicing requests for data 15. In this manner, the central proxy server 12 acts as a sort of "Nielson Ratings Service" for the Internet. This information may be provided either free or for a fee to interested parties.

A heuristic determination module 176 utilizes the statistical information gathered by modules 172 and 174 in making the heuristic determination of whether to keep a particular file of data 15. If the determination is to keep the data 15, a least recently used (LRU) module may be used to decide which existing data is to be discarded to make room for the new data 15. The LRU calculation module 178 calculates the amount of use of each file of data 15 and adds a local use factor into the determination. This heuristic determination is applied by the heuristics application module 166 as described above.

The heuristics determination module 176 preferably utilizes localized web page demand statistics compiled by the local proxy server 22 as well as globalized web page demand statistics collected by and transmitted periodically from the central proxy server through satellite transmission.

A hit rate reporting module 179 periodically reports the local hit rates for part or all of the data 15 transmitted from the central proxy server 12. This information is used to calculate global demand statistics. Preferably, each participating local proxy server 22 tabulates every request from every end user station 30. A hit report is then periodically transmitted to the central proxy server 12. The central proxy server 12 in turn tabulates the accumulated hit reports for all web pages requested and periodically sends updated usage information with the hit totals to the local proxy servers 22. Preferably, every file of data 15 is assigned a globally unique number which can be stored and transmitted compactly and which is used to identify the particular data 15.

proxy server 22 services a research institution, the policy 204 may give high weight to data 15 relating to the research being conducted highly, etc.

Thus, as shown in a global demand calculation block 180 of Figure 3, a global demand monitoring module 182 resident within the central proxy server 12 monitors all of the requests for information from each local proxy server 22. It also monitors all of the hit rate data transmitted by the reporting modules 179 of each local proxy server 22. In response, a global demand compilation module compiles the data collected by the global demand monitoring module 182. Within those statistics may be information relevant to each particular subject matter such that the heuristics procedures may take the categorized demand data into account according to the particular weighting of the subject categories in their local policies 204.

A particular subscribing local proxy server 22 might be an educational institution, business, or news provider, and might weight subject matter statistics from relevant categories more heavily within its own local policy 204 than others of the local proxy servers 22. A global demand transmission module 186 transmits the demand data, preferably by satellite transmission 34, in the same manner as the requested data 15 is transmitted.

Figure 6 is a schematic flow chart diagram illustrating one manner of operation of a software program 220, a copy of which preferably operates on each subscribing local proxy server 22. In one embodiment, the software program 220 comprises the local caching module 17. The program 220 initializes at a start block 222. The program 220 then proceeds on to a block 224 where it awaits receipt of input to be processed. If the input is a terminate command, the program 220 receives the command at a block 226 and then progresses to an end block 228, where the program 220 terminates.

Three representative procedures are shown in Figure 6, corresponding to several different general functions that may be accomplished by the program 220. For instance, if the program 220 receives a user request for data 15, the program receives the request at a block 230. The program 220 then proceeds to a block 232, where the caching protocol of the http language is utilized to consult the local cache database 24 and thereby determine if the requested data 15 is present. At a query block 234, the program 220 branches in one of two directions depending on whether the data 15 is present in the local cache database 24. If the data is present, the program proceeds to a block 236 where it transmits the data 15 to the requesting station 30 for presentation to the user. The program then proceeds to a block 238, which returns control of the program to the start block 222.

data to make a heuristics determination at a block 272. The program 220 then proceeds to a block 274 which passes control back to the receive input block 224.

Figure 7 is a flow chart diagram illustrating one manner of operation of a software program 300 resident within and operating on the central proxy server 12 of Figures 1 and 2. The program 300 initializes at a start bock 302 and proceeds to a block 304, where it awaits receipt of a request for data 15. Of course, the program 300 could have other means of identifying high demand data 15. Web site search engines or other pre-generated statistical data could be consulted, for instance. The central proxy server 12 could also keep a history of data requests for web pages 15a, which are updated daily or weekly, etc. in order to re-transmit those pages 15a once updated.

Once data 15 is requested by a local proxy server 22 or otherwise identified as being of sufficient demand to merit caching within the subscribing local proxy servers' attendant cache databases 24, the program 300 proceeds to a block 305. Here, the program 300 requests a copy of the data 15 from the hosting Internet site 35. Subsequently, as designated by a block 306, the data 15 is received over the Internet 20 from the hosting site 35. The data 15 may then be filtered before being transmitted and/or stored within the master cache database 312. For instance, a morality filter could be used to filter out obscene language or pornographic materials. Other types of filters could also be used by content, or filters could be used to filter out data for which it is known that hit rates in the future will be low. A heuristic scheme or some type of request history might be employed for so doing.

At a block 310, the data 15 which has successfully passed through filtering is stored in the master cache database 14 as represented at a block 210. The data 15 is then sent to the broadcast uplink transmitter 16 as represented by a block 312. As discussed, the communication medium may be any suitable medium, but satellite transmission will be discussed herein. As shown by a block 314, the data 15 is then transmitted to the satellite 18. Thereafter, as shown by a block 316, the data 15 is transmitted by satellite transmission 34 to the subscribing local proxy servers 22. This transmission may be coordinated by the program 300.

At a subsequent block 318, global demand statistics are updated to reflect the request for the data 15. The updated global demand statistics may be transmitted at this point to the local proxy servers, or this information may be periodically transmitted. Subsequently, the program 300 progresses to a block 322 which returns control to the start block 302.

352 is also preferably provided and handles all user interface types of interaction. The central proxy server 12 is preferably in communication with a database 334 which may employ a schema for handling historical logging of information transmitted back from the local proxy servers 22 to the database 334 upon a message that is sent by the central proxy server 12.

The web data 15 which may be handled and forwarded through the system include web objects or groups of web objects that form a web page 15a. Web objects are also referred to as HTML objects or HTML data.

When the data 15 is placed on the priority queue, the priority of that data is altered somewhat continuously as the central proxy server 12 receives indication from the local proxy servers 22 that one or more end users have made a cache hit on that object or that a local proxy server 22 reported another miss. This popularity information keeps accumulating even after the initial insertion into the priority queue and the highest priority data files 15 are considered for transmission.

The transmission queue 346 draws the highest priority data files 15 off the priority queue and then attempts to gather together the web objects, container objects, and their component objects and form that data 15 into logical messages. The input side of the transmission queue 346 receives the data 15. The output of the transmission queue 346 is a list or multiple lists of logical messages that represent pieces of that data 15.

The transmission queue 346 processes the data 15 and formulates logical messages, all of which are preferably small, such as about 8 Kbytes. These logical messages are placed onto an output queue, or multiple output queues, and that data serves as input to the packetizer module 348.

The responsibility of the packetizer module 348 is to take those logical messages, whatever size they are, and make them fit into IP multicast packets in such a way that there is no wasted space. In other words, if there are several small logical messages, they would get packed one end-to-end inside of those IP multicast packets.

The maximum packet size is preferably selected to stay within the maximum data payload of an Internet frame, about 1500 bytes. But more importantly, the length of the data in each of the packets is preferably optimized to be approximately 1442 bytes. That number is selected in one embodiment to work with the subsequent transmission of that packet over a transmission medium such as a satellite system which incorporates DVB packets and the multi-protocol encapsulation or MPE encodings. The size is chosen to be optimal for data

Each data file 15 is preferably broken up into multiple object messages 360 from 1 to N. Each object message 360 consists of an object message header 362 and a payload 364. Each object message 360 is then broken up into a satellite packet 366 by the packetizer module 348. The Satellite packets, as stated, are 1442 bytes in size or may be another suitable size selected to give the optimal satellite throughput.

The satellite packet similarly consists of a packet header 368 and a payload 370. There are 1 to M packets for every given object message. The satellite packets are then converted into IP multicast format via industry standard operations. The IP multicast is then converted into a digital video broadcast (DVB) stream, which is once again the industry standard.

The DVB packets can be constructed using a hardware and software solution. An example of this is a DVB packetizer product which is provided by Skystream International. On the receive side, the DVB packets are received by a satellite modem, which then rebuild those packets into IP multicast format. The thusly formatted packets are then received by the local proxy server 22, reassembled, and eventually the entire web object 15 is reconstructed using the inverse of the process previously described.

Example of Local Proxy Server Operation

Referring now to Figure 10, shown therein is a block diagram illustrating the operation of one embodiment of a local caching module 17 of a local proxy server 22 of Figure 1. Within Figure 10 is shown a satellite receiver 372. Emanating from the satellite receiver 372 are transmission signals 374 and 376 indicating network transmissions using multicast packets for Internet data files 15. The data files 15 are shown being received by a receive assembler module 378.

Each depicted module represents a software subsystem or component of the local caching module 17. A receive assembler 378 receives the IP multicast packets and as it receives one packet and then the next packet and so on, as fast as it is able to, it reconstructs the object messages, which represent portions of data files 15.

The data files 15 are transmitted via an message dispatcher 380 , which is basically a switching mechanism, to an injector module 392. The injector module 392 accumulates the object messages, one or more object messages per web object, reassembles those messages into the web object, and injects that web object data 15 into the local cache database management module 29.

database management module will call up the miss manager subsystem 400 and identify the URL (universal resource locator) of that web object on the Internet and give the local caching module 17 an opportunity to take steps to find and gather that web object data and deliver it to the local cache database management module 29 by means of the present system, which is what has been described up to this point, where the data is received over the satellite by the local caching module 17 from the Central proxy server and injected or discarded back into the local cache database management module. So the local cache database management module is given first right of refusal and will try to satisfy the end user request for a particular web option.

Example of Hit and Miss Rate Reporting

Referring to Figure 10, the local cache database management module 29 is preferably configured to report cache misses to the miss manager 400 of the local caching module 17. In response, the miss manager 400 preferably then transmits a message immediately to the central proxy server 12 via the message dispatcher 380 and the central proxy server session manager 382 to instruct the central proxy server 12 that the local proxy server 22 does not contain a copy of the web object 15 the end user is looking for.

The central proxy server 12 then preferably uses the miss information in calculating global miss rate information. That is, the central proxy server 12 uses the miss reports that it receives from the local proxy servers 22 in the overall system 10 to compute the relative priority of web objects 15 in the priority queue 342.

Referring back to the local caching module 17 of Figure 10, also provided therein in the depicted embodiment is a hit manager module 396. The hit manager module 396 is similar to the miss manager 400 in that when a client requests a web object 15 from a local cache database management module 29 and the cache database management module 29 has that object already stored in its cache, a hit is registered. The local proxy server 17 accumulates hit counts on an object-by-object basis as they are reported. It then reports that information at least every few seconds to the central proxy server 12 advisor via the message dispatcher and the central proxy server session manager.

The local caching module 17 preferably makes those hit reports under several different circumstances. Initially, it may only have room to track hits on so many objects at once, and that determines somewhat the size of the report messages that it is going to send to the central

that object and a new priority for that object can be calculated using miscounts and hit counts to date and some other information.

Every time a hit or miss report is received by the central proxy server 12 , it will accumulate that and update the totals, but it may not actually transmit, or may not complete the recalculation of the priority because doing so requires removing the object from the priority queue and reinserting it at a new position. Given that there will likely be hundreds of thousands of web objects represented from that priority queue, that becomes an expensive proposition in terms of computer processing time. So instead the potential change in the priority are evaluated and when it crosses a certain threshold or every so many times a report has been received, then the object is reinserted into its appropriate position.

Relative priorities of any particular object are preferably transmitted under two circumstances. When the object data itself is transmitted, the priority of that object is transmitted with it. That information is provided to the local caching module 17 and to the local cache database management module. The local cache database management module at that point preferably incorporates that global hit data that has been provided into its own internal calculations of how wonderful this object is and how long it wants to keep it around in its local cache. The second time that n object's priority is typically transmitted is when that priority has changed noticeably, somewhat substantially. But the object data itself would not change. There would be no need to resend the subject data but it is useful to send the very brief, very small administrator message saying the object identified by global ID such and such has a new global priority, so please consider it.

In addition to recording hit counts on a moment to moment basis for various objects, the hit manager preferably also keeps the longer term total of those hits. In fact, it preferably keeps a total for a time period of, for example, fifteen minutes. Every object that has had any hits on it in 15 minutes, at the end of that 15 minutes is then evaluated relative to all the others. The objects are sorted out and the record of the ordered top 100, 200 or some selected number of objects are logged to a file in the local cache box and that file will be available for subsequent retrieval by the central caching management module 13 and the historical database agent

The DB agent goes out periodically and queries each of these local caching modules 17 getting all of the accumulated 15 minute logs and data regarding what was most popular of your local cache during that 15 minute, and that will update all the different local caching modules 17 and will present a number of opportunities for data mining.

of those functions is called with a short command, and if necessary, accompanying parameters. Those commands are responded to with the performance of the requested function, which may, for instance, include the depositing of data in a selected buffer or memory location.

As an additional example, when the injector module 392 receives web objects or other files 15 and needs to transmit those objects to the cache database management module 29, it uses commands from selected API instruction sets and passes those commands to the cache database management module 29 to achieve that desired function.

Other API functions may include an instruction or command which requests the cache database management module 29 to locate an object 15 within the local cache database 24. Another API instruction may notify the cache database management module 29 that a particular data file 15 is being sent and is to be saved together with another stored data file 15. This data may not actually be part of the original object data 15, but may be related information, such as a component of a web page 15a.

In one embodiment, the cache database management module 29 may respond to the instruction asynchronously, that is, at varying times. The responses may be immediate, or may be delayed, to acknowledge the results of an initial request, for instance. Other functions may comprise cache database management module one-way communication to the local caching module 17 to provide notification of certain events such as a hit report or a miss report.

Thus, an API interface is very highly integrated between the local caching module 17 and the cache database management module 29. This provides the advantages that the cache database management module 29 can be communicated with internally and unique types of information can be received back from the cache database management module, including statistics such as hit reports, almost instantaneously.

Integration does require some additional support and alteration by the vendor and produces the results of greater flexibility of management and information extraction. Additionally, speed is increased, so that reported information is more current. In prior art arrangements, where an software applications residing with different external servers communicate, there are necessarily slowdowns resulting from network transmission delays, processing of network messages in and out of the boxes, and slowdowns resulting from hardware bottlenecks. These include the use of drivers of the hardware and slow funneling of information that must work its way up through the various hardware protocols and then back

Transmission of Web Pages in Their Entirety

The system 10 of the present invention allows a web page to be transmitted in its entirety, rather than transmitting a listing of the contents of the page and going back separately requesting those contents as is conducted with most Internet accesses. One feature preferably provided within the cache database management module 29 of the present invention is the ability to examine an HTML web object (such as a web page) 15 and determine its component parts. The HTML object preferably includes a container that describes the complete layout of a particular page 15a that will be seen on the browser window 25 of a user 30. The container or other such directories typically contain most of the readable text and contain specific instructions, using the hypertext mark up language syntax, to identify other web objects, images, Java applets, or other types of web object that are part of the page.

The cache database management module also preferably looks at a potential container that refers to all of the other images and elements that comprise the total Web page experience. The cache database management module 29 will, upon receiving the HTML object 15, scan through it immediately, parsing out the particular references to other objects that are components of that page. As the cache database management module 29 finds each of those, it will initiate a request. The system 10 first checks to see if the image is already present in the local cache database 24. If it is not, the system 10 initiates a request over the Internet 20 to retrieve that image and continues to do that for every object that was indicated as a component of that HTML container object.

This feature provides an additional acceleration to the end user 30. The cache database management module 29 has hopefully retrieved some, if not all, of the component objects that complete a web page by the time that the end user browser 25 has received the original HTML content. The cache database management module 29 preferably begins to analyze and parse the requested object 15 and goes back to the cache database management module 29 requests all component objects of the web page 15. The entire contents of a web page 15a are then streamed to the web browser 25, without waiting for the web browser 25 to request them individually. Without this read-ahead feature ,as it is called, on the cache database management module 29, the browser 25 after receiving the original HTML object, would then have to request each of the component objects with a separate request, and each of those may result in cache request, and subsequently, a request across the Internet. Whereas with the read ahead feature of the cache database management module, hopefully many of those objects 15 are already in the cache database management module and the client browser 25 will receive

CLAIMS:

1. A system for accelerating distribution of information over a global communications network, the system comprising:

a central proxy server configured to extract data files from a global communications network and transmit the extracted data files over a communication medium;

a local proxy server configured to receive the extracted data files from the central proxy server over the communication medium, provide the extracted data files to a caching database in local communication with the local proxy server, and communicate with a plurality of customer stations to receive requests for information available at remote locations across the global communications network;

a cache database management module configured to communicate with the caching database in order to store the extracted data files within the caching database and to retrieve the extracted data files from the caching database; and

a local caching module operating within the local proxy server and in integral communication with the cache database management module, the local caching module configured to receive the extracted data files from the communication medium and instruct the cache database management module to store at least a portion of the extracted data files within the caching database,

the local proxy server also configured to respond to the requests from the customer stations by the cache database management module checking whether the requested information is among the extracted data files stored within the caching database, and if the information is not among the extracted data files stored within the caching database integrally communicating with the local caching module that the information is not present.

2. The system of claim 1, further comprising an application program interface (API), the local caching module in integral communication with the cache database management module through the API.

3. The system of claim 1, wherein the cache database management module is a commercially available Internet caching system program.

4. The system of claim 1, wherein the local caching module is configured to, through the integral communication with the cache database management module, receive

determination whether to retain the data files, the policy being unique to the local proxy server.

14. The system of claim 12, wherein the local proxy server is configured to receive from the central proxy server global statistical usage data regarding data files previously transmitted from the central proxy server to the local proxy server.

15. The system of claim 14, wherein the heuristic procedure includes consideration of local statistical usage data regarding the frequency of requests for specific data files together with consideration of the global statistical usage data.

16. The system of claim 12, wherein the heuristic procedure accounts for a particularized subject matter of interest to users of the local proxy server in making the determination whether to retain the data files, and wherein the global statistical usage data includes data indicative of the demand for data files specific to the particularized subject matter.

17. The system of claim 1, wherein the central proxy server is configured to transmit extracted data files to a plurality of local proxy servers, and wherein private data files are transmitted in addition to global data files, the private data files being received by subscribing local proxy servers to a private network and not received by the others of the plurality of local proxy servers which are not subscribers to the private network.

18. The system of claim 1, wherein the central proxy server is further configured to assign all extracted data pages a unique identifier code and transmit the unique identifier codes over the communication medium together with the extracted data pages.

19. The system of claim 18, wherein the central proxy server is further configured to ascertain a popularity rating of each extracted data file and correlate a popularity rating of each extracted data file with the unique identifier code for that extracted data file.

20. The system of claim 18, wherein at least a portion of the extracted data files comprise web pages having multiple components normally transmitted separately, and further comprising associating the multiple components of a web page together such that when a customer station requests the web page, the multiple components of the web page are transmitted to the customer together.

21. A method for accelerating distribution of information over a global communication network, the method comprising:

extracting data files from a global communications network;

transmitting the extracted data files over a communication medium;

29. The method of claim 21, further comprising responding to requests from a customer station for information available at a remote location on the global communication network by determining whether the information is present in the caching database and if the information is present, transmitting the selected data file to the customer station and if the information is not present, transmitting a notification to a central proxy server that the information was requested by a customer station.

30. The method of claim 21, further comprising operating the local caching module primarily in an IPX protocol native to the local proxy server.

31. The method of claim 21, further comprising receiving into the central proxy server updates from a plurality of communicating local proxy servers regarding the number of requests for data files residing at remote locations on the global communication network, compiling the updates to determine the relative demand for the data files and transmitting the most popular data files concurrently to a plurality of communicating local proxy servers over the communication medium.

32. The method claim 21, further comprising conducting a heuristic determination of whether to retain the extracted data files in the caching database or to discard the extracted data files.

33. The method of claim 32, wherein conducting a heuristic determination further comprises assigning a policy of different weights to different types of extracted data files, the policy unique to the local proxy server.

34. The method of claim 32, further comprising receiving from a central proxy server global statistical usage data regarding data files previously transmitted from the central proxy server to the local proxy server.

35. The method of claim 34, wherein conducting a heuristic determination further comprises considering local statistical usage data regarding the frequency of requests for specific data files together with considering the global statistical usage data.

36. The method of claim 32, wherein conducting a heuristic determination further comprises accounting for a particularized subject matter of interest to users of the local proxy server in making the determination whether to retain the data files, and wherein receiving the global statistical usage data includes receiving data indicative of the demand for data files specific to the particularized subject matter.

37. The method of claim 21, further comprising transmitting extracted data files to a plurality of local proxy servers over the communication medium and transmitting, in

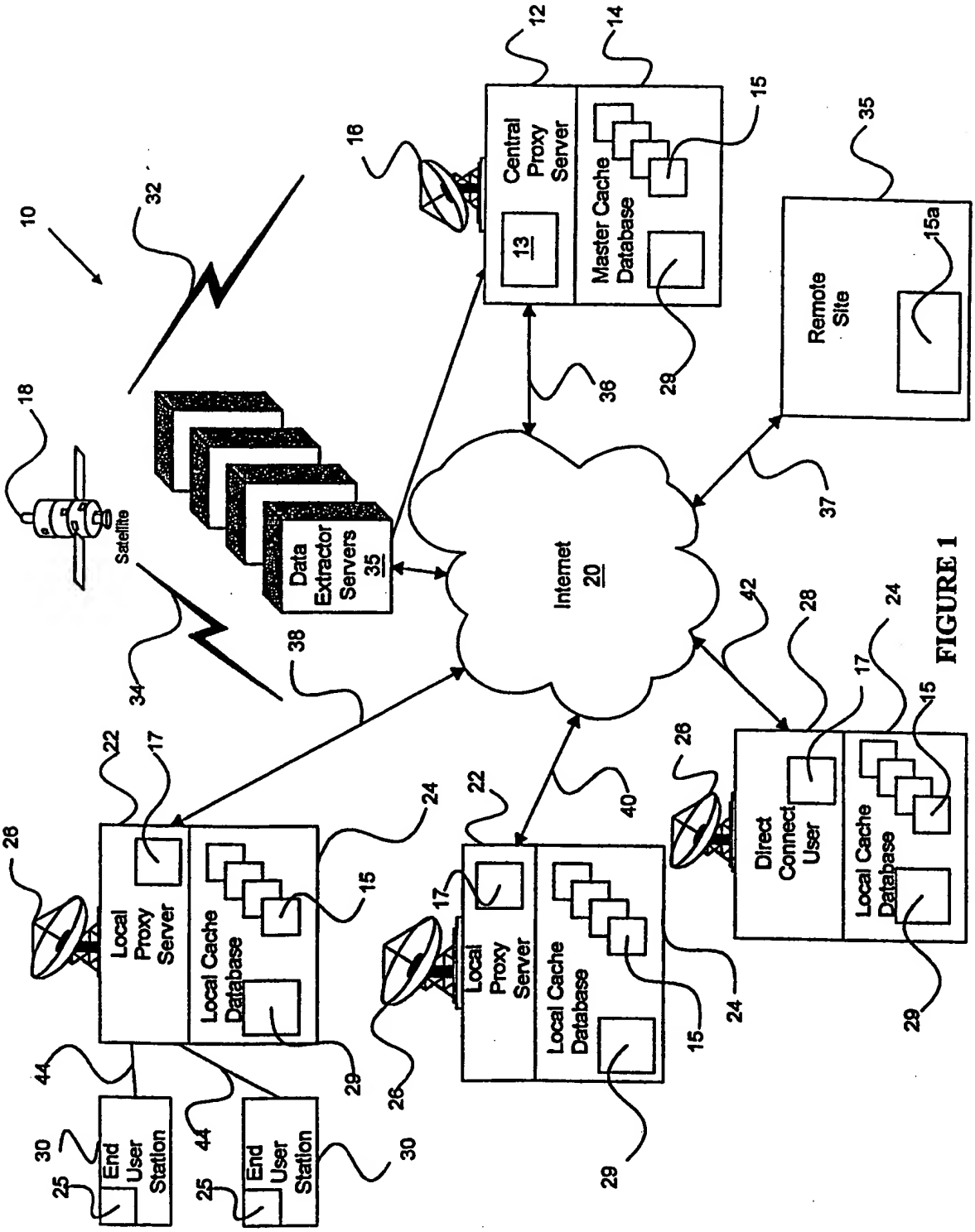

10
32
16
Central Proxy Server
13
Master Cache Database
12
14
15
35
Remote Site
15a
18
Satellite
34
Data Extractor Servers 35
Internet 20
36
28
37
38
42
40
Local Proxy Server
Local Cache Database
22
17
24
26
44
30
25
End User Station
Direct Connect User
29

FIGURE 1

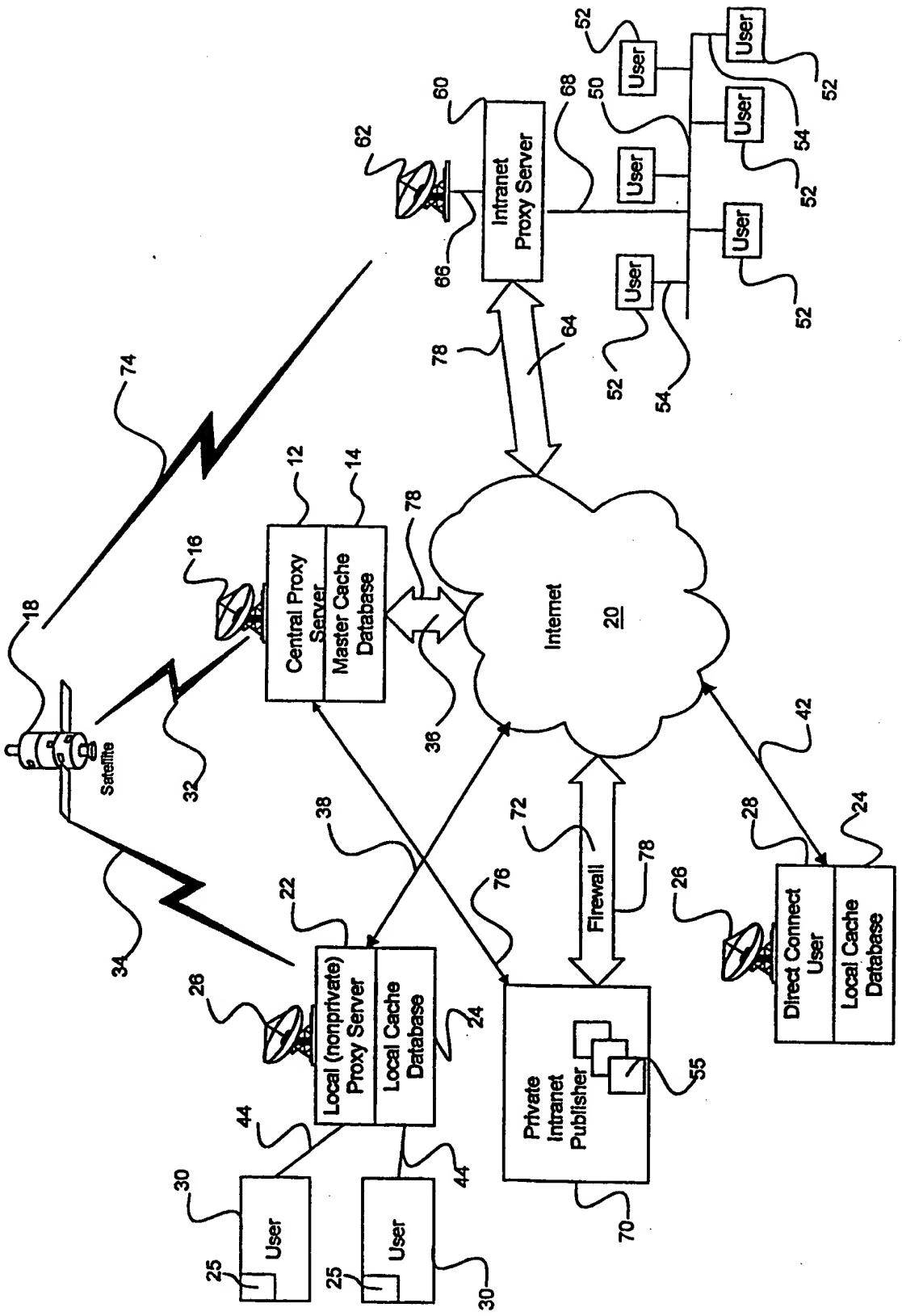

Satellite
Central Proxy Server
Master Cache Database
Internet
20
Intranet Proxy Server
Local (nonprivate) Proxy Server
Local Cache Database
Private Intranet Publisher
Firewall
Direct Connect User
Local Cache Database
User
12
14
16
18
22
24
25
26
28
30
32
34
36
38
42
44
50
52
54
55
60
62
64
66
68
70
72
74
76
78

FIGURE 2

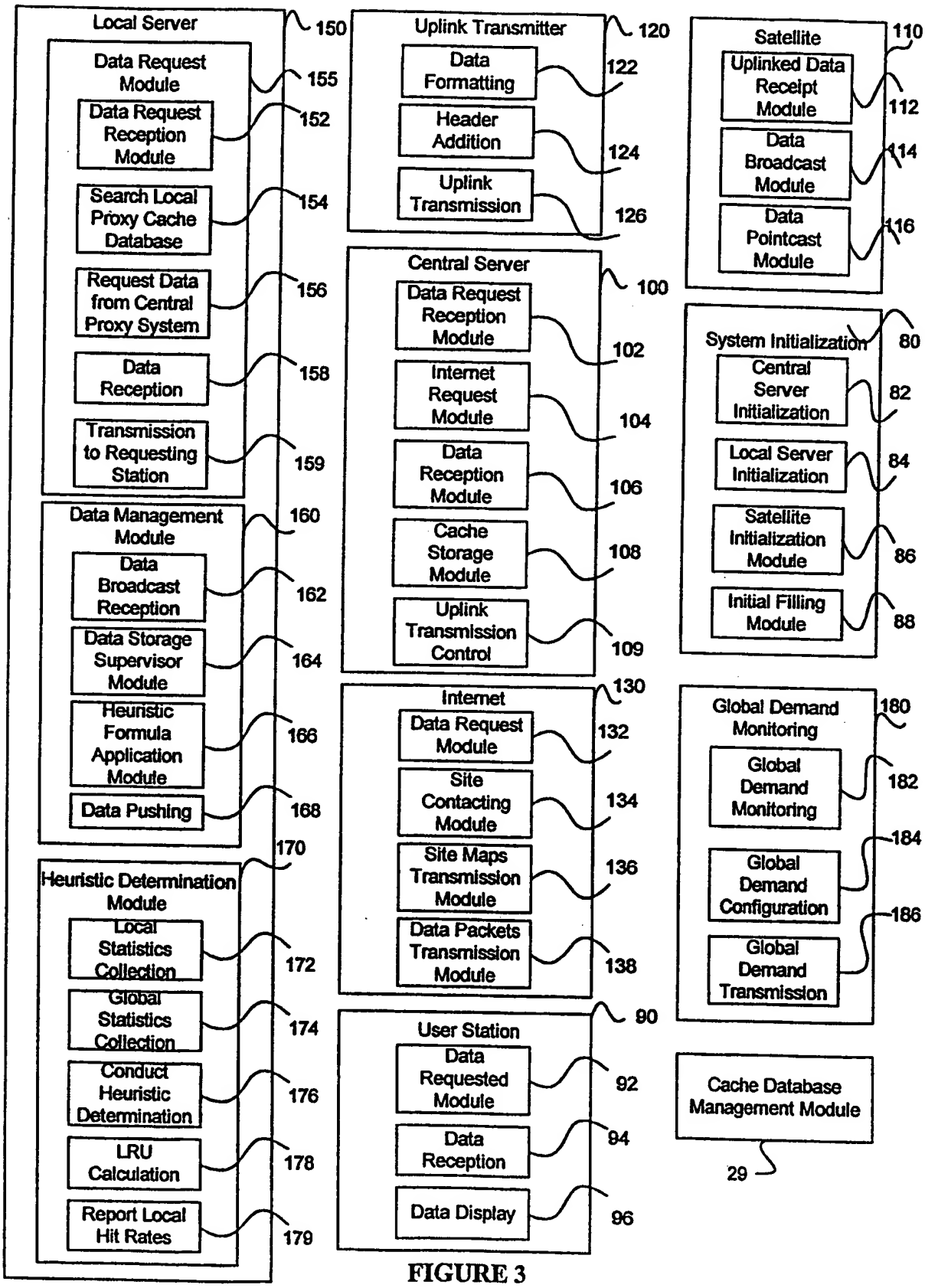

Local Server
150
Data Request Module
155
Data Request Reception Module
152
Search Local Proxy Cache Database
154
Request Data from Central Proxy System
156
Data Reception
158
Transmission to Requesting Station
159
Data Management Module
160
Data Broadcast Reception
162
Data Storage Supervisor Module
164
Heuristic Formula Application Module
166
Data Pushing
168
170
Heuristic Determination Module
Local Statistics Collection
172
Global Statistics Collection
174
Conduct Heuristic Determination
176
LRU Calculation
178
Report Local Hit Rates
179
Uplink Transmitter
120
Data Formatting
122
Header Addition
124
Uplink Transmission
126
Central Server
100
Data Request Reception Module
102
Internet Request Module
104
Data Reception Module
106
Cache Storage Module
108
Uplink Transmission Control
109
Internet
130
Data Request Module
132
Site Contacting Module
134
Site Maps Transmission Module
136
Data Packets Transmission Module
138
User Station
90
Data Requested Module
92
Data Reception
94
Data Display
96
Satellite
110
Uplinked Data Receipt Module
112
Data Broadcast Module
114
Data Pointcast Module
116
System Initialization
80
Central Server Initialization
82
Local Server Initialization
84
Satellite Initialization Module
86
Initial Filling Module
88
Global Demand Monitoring
180
Global Demand Monitoring
182
Global Demand Configuration
184
Global Demand Transmission
186
Cache Database Management Module
29

FIGURE 3

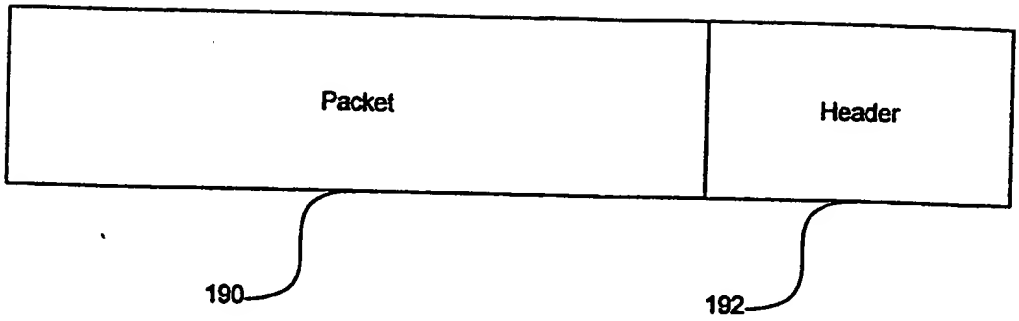


FIGURE 4

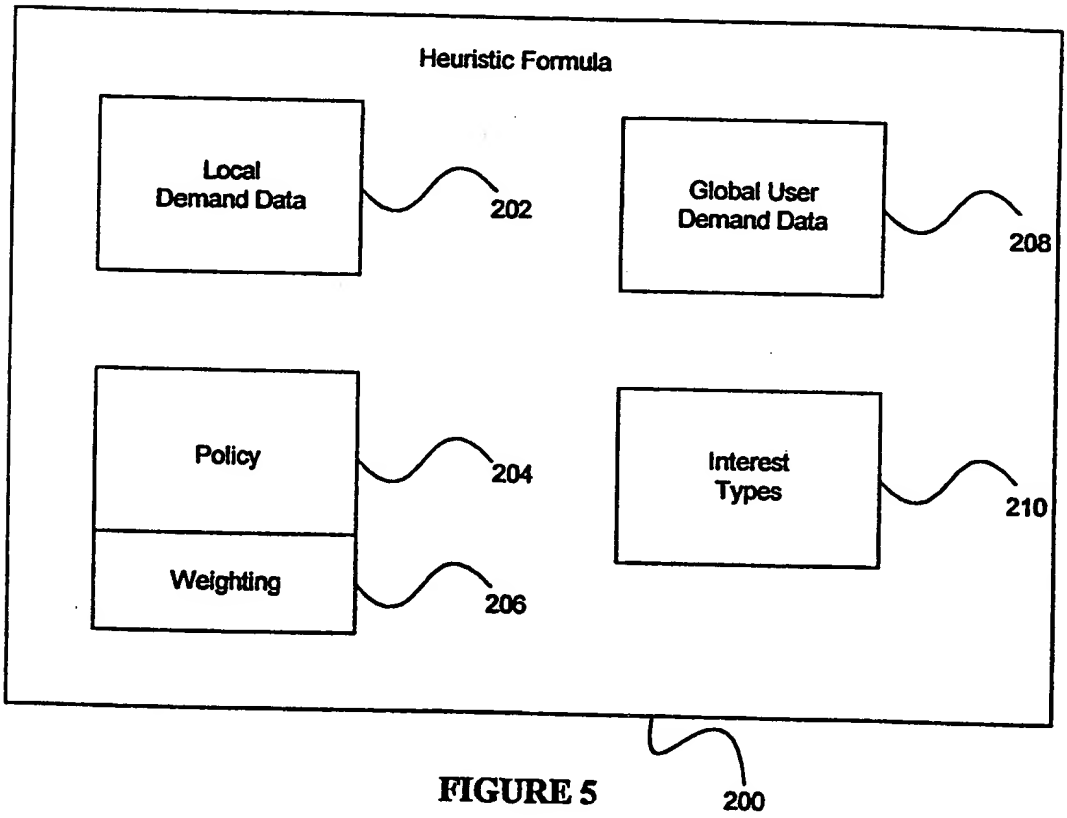


FIGURE 5

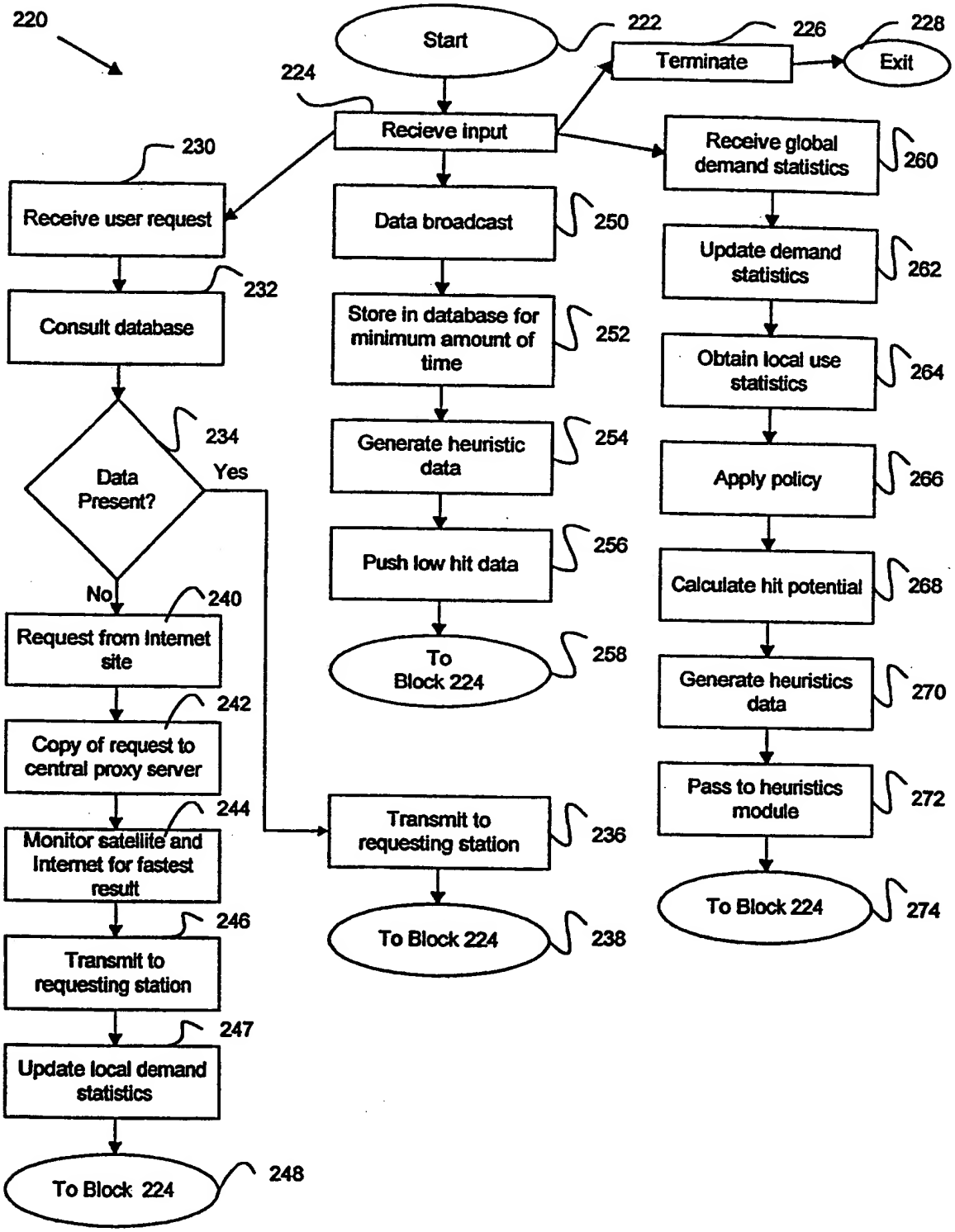

220
Start
222
Terminate
226
Exit
228
224
Recieve input
230
Receive user request
Consult database
232
Data Present?
234
Yes
No
240
Request from Internet site
242
Copy of request to central proxy server
244
Monitor satellite and Internet for fastest result
246
Transmit to requesting station
247
Update local demand statistics
To Block 224
248
Data broadcast
250
Store in database for minimum amount of time
252
Generate heuristic data
254
Push low hit data
256
To Block 224
258
Transmit to requesting station
236
To Block 224
238
Receive global demand statistics
260
Update demand statistics
262
Obtain local use statistics
264
Apply policy
266
Calculate hit potential
268
Generate heuristics data
270
Pass to heuristics module
272
To Block 224
274

FIGURE 6

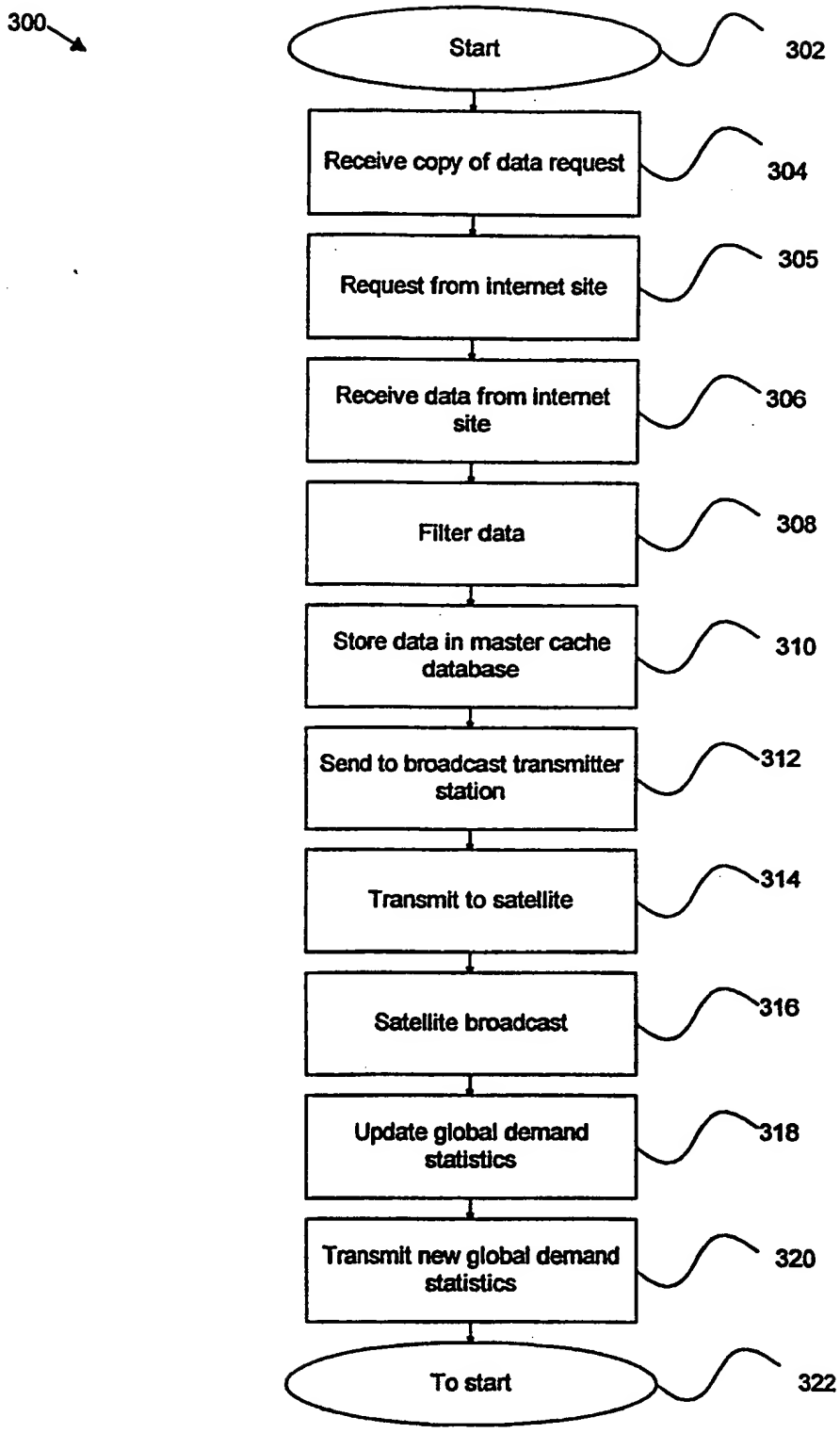

300
Start
302
Receive copy of data request
304
Request from internet site
305
Receive data from internet site
306
Filter data
308
Store data in master cache database
310
Send to broadcast transmitter station
312
Transmit to satellite
314
Satellite broadcast
316
Update global demand statistics
318
Transmit new global demand statistics
320
To start
322

FIGURE 7

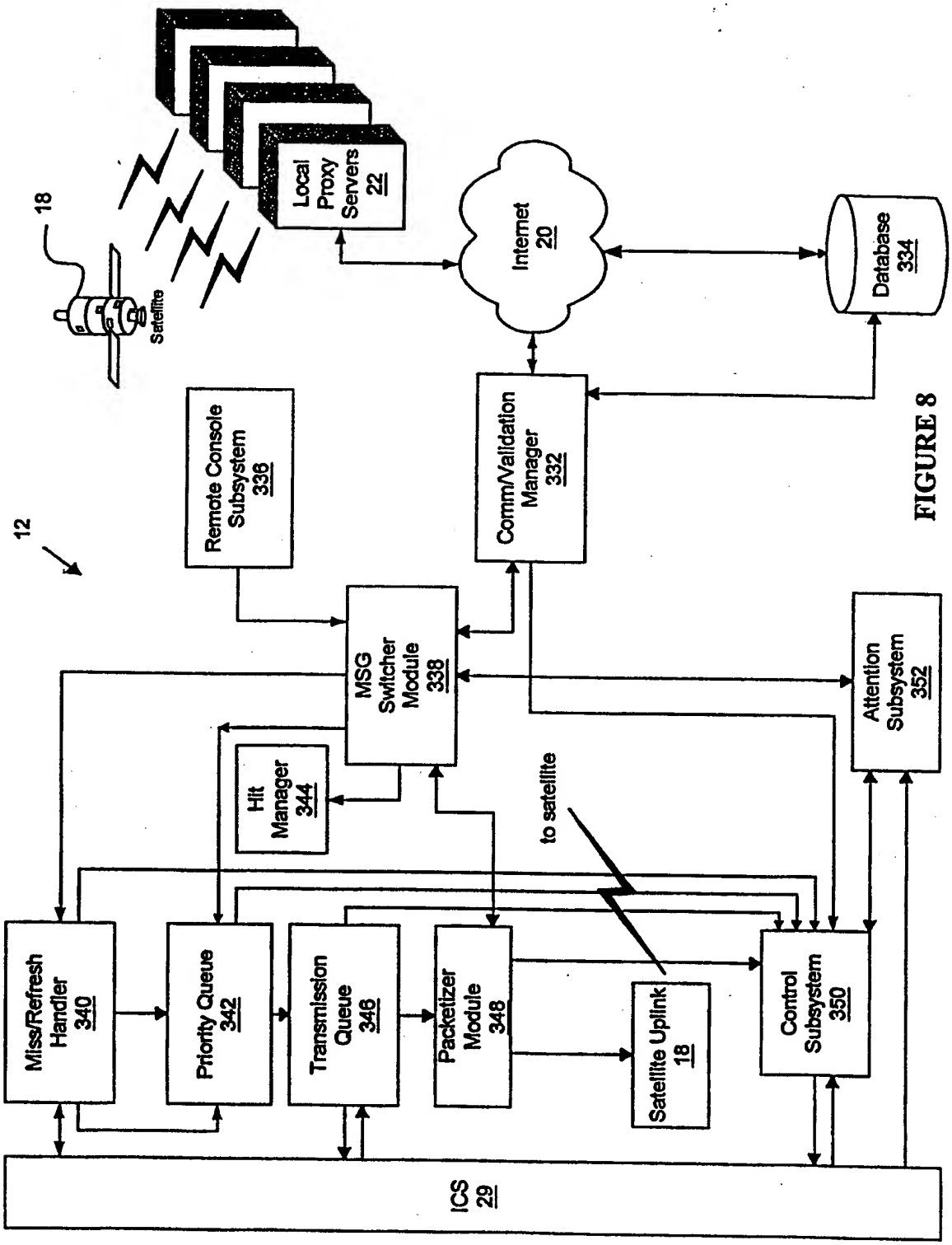

18
Satellite
Local Proxy Servers 22
Internet 20
Database 334
12
Remote Console Subsystem 336
Comm/Validation Manager 332
MSG Switcher Module 338
Attention Subsystem 352
Hit Manager 344
to satellite
Miss/Refresh Handler 340
Priority Queue 342
Transmission Queue 346
Packetizer Module 348
Satellite Uplink 18
Control Subsystem 350
ICS 29

FIGURE 8

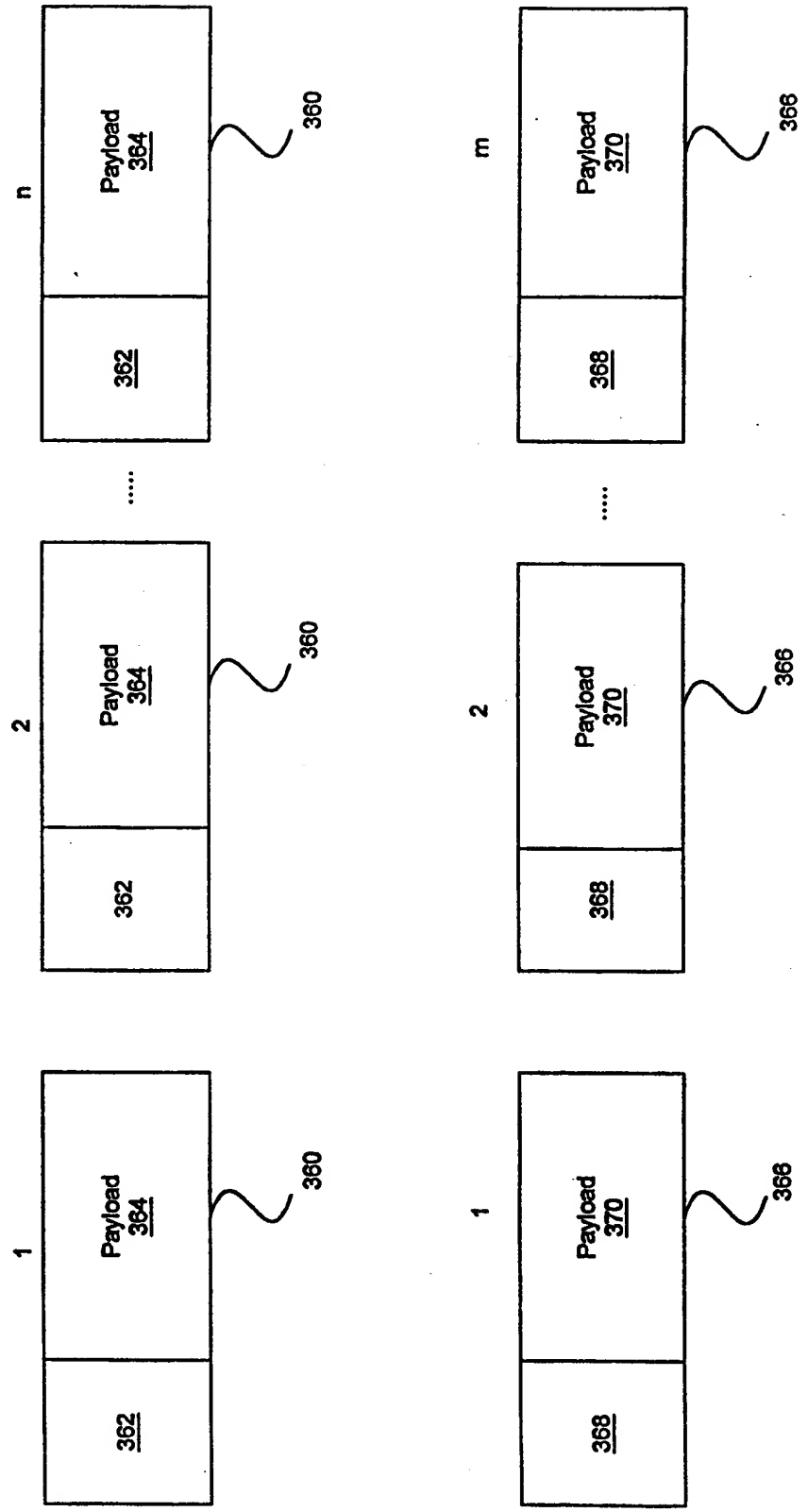


FIGURE 9

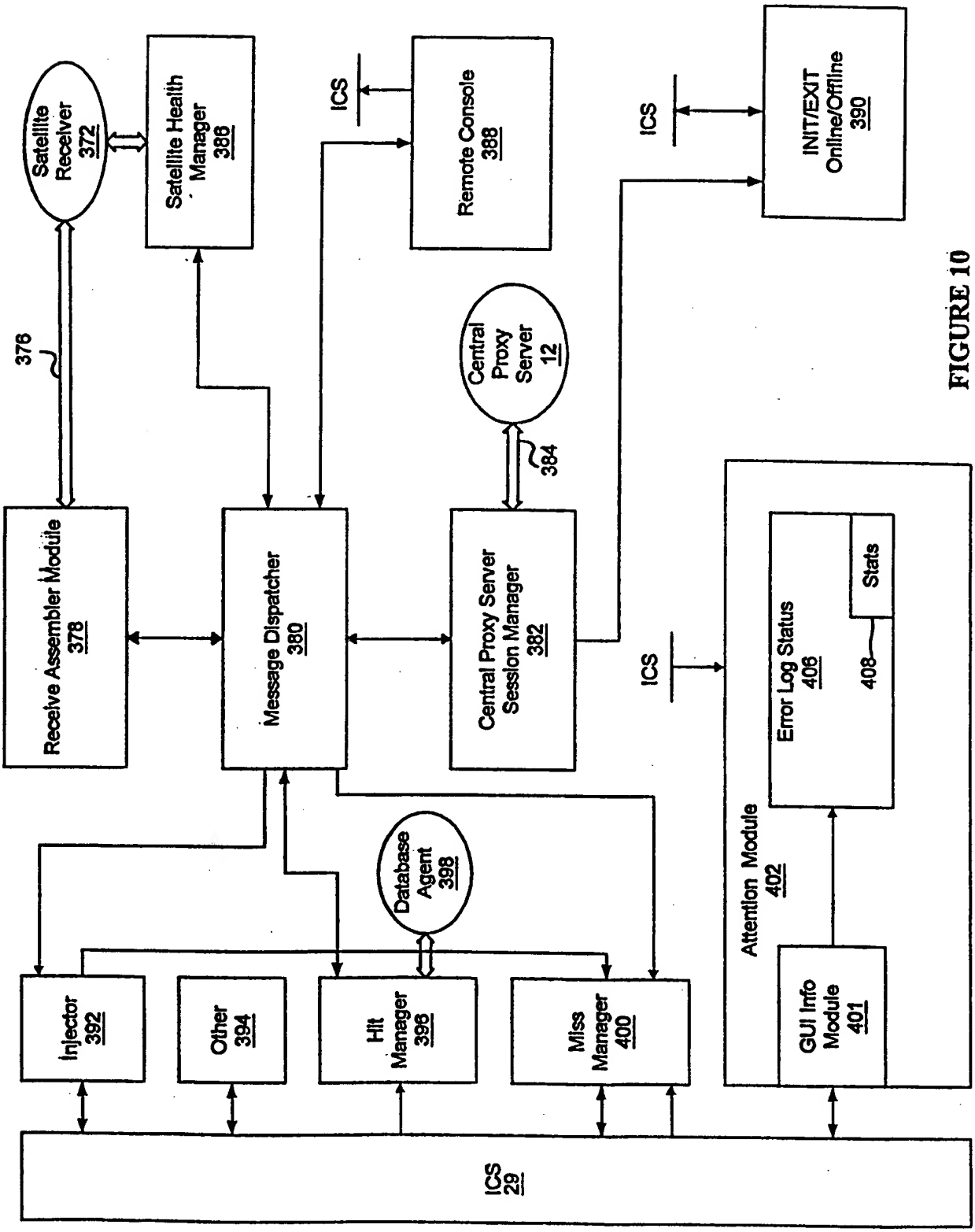

Satellite Receiver 372
Satellite Health Manager 386
ICS
Remote Console 388
ICS
INIT/EXIT Online/Offline 390
376
Central Proxy Server 12
384
Receive Assembler Module 378
Message Dispatcher 380
Central Proxy Server Session Manager 382
ICS
Attention Module 402
Error Log Status 406
Stats
408
Database Agent 398
Injector 392
Other 394
Hit Manager 396
Miss Manager 400
GUI Info Module 401
ICS 29

FIGURE 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/00587

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :G06F 15/16
US CL :709/217, 235

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 709/217, 235

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

None

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WEST, EAST, IEEE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y,P	US 6,012,085 A (YOHE et al.) 04 January 2000, col. 5-12	1-40
Y,P	US 6,003,082 A (GAMPPER et al.) 14 December 1999, col. 4-12	1-40

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 18 MAY 2000	Date of mailing of the international search report 12 JUN 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer AHMAD MATAR For Eugenia Zogyan Telephone No. (703) 305-4731

Form PCT/ISA/210 (second sheet) (July 1998) ★